

Individual Assignment 4: Maps

Chris Anderson
PREDICT 455 Section 56

Summary and Problem Definition

A professional baseball franchise seeks to optimize its scouting department by deploying scouts in the most efficient manner possible. Typically, scouts are assigned a particular region of the country, or the world, and focus their efforts on finding high school and college players in that geographical area. This exercise will focus on the continental United States and use data visualization to examine the 2016 MLB draft, specifically the number of draft picks from each state.

Methods

A table containing the results of the 2016 MLB draft was obtained from Baseball America's draft database and saved in .csv format. This file contained the names of each of the roughly 1200 players drafted, along with a few additional details for each – including the two letter abbreviation of the state from which they hailed. In addition, a list of state names and abbreviations was downloaded and saved in .csv format (Softschool.com). These files were merged in order to use the mapping functionality within the statistical software R, which relies on state names rather than abbreviations. The number of draft picks from each state was calculated and used to visualize the data.

Programming Overview

In R, a number of packages, including *sp*, *maptools*, *ggplot2*, *maps*, and *mapproj*, were used to analyze the data, create geographical shapes, and plot a choropleth map. Both .csv files were loaded, the count of players drafted was calculated per state, and the tables were merged to add the state name to the frequency table. This table provided the data for the map visualization. Next, a file containing data required to create geographical shapes within R was loaded from www.gadm.org. After a few data preparation and validation tasks, the shape data was merged with the frequency table. A map of the

continental United States was produced with shading indicating the number of draft picks from each state. The map was exported to both .pdf and .svg format.

Results

The resulting graphic reveals patterns in the data that may prove beneficial to a professional baseball team looking to optimize its scouting department. California produced the most draft picks, by far, so it may be wise to allocate at least one scout to the state on a full-time basis. Washington, Oregon, Nevada, and Arizona also produced a fair number of draft picks. This “west-coast” region should receive a significant portion of the available scouting resources. Predictably, many other southern states along the gulf and southeastern coasts generated draft picks in larger volumes. More scouts should be sent to cover these areas, as well. The map also highlights regions from which only a few players were selected, if any. It may be possible for a single scout to cover the entire mountain west region of Idaho, Montana, Wyoming, Utah, and the Dakotas. It may be tempting to suggest that these areas should be disregarded completely, but this would be unwise. While the odds of finding a player may be slim, a team must explore all possible avenues or risk missing out on a potential star. There are, of course, many other factors to be considered when deploying baseball scouts, including travel, prospect schedules, and cost. However, this exercise may be a useful resource when deciding where to allocate these important resources.

References

2016 Draft Database. (2016). Retrieved November 06, 2016, from <http://www.baseballamerica.com/draftdb/index.php>

Softschools.com. (n.d.). Retrieved November 06, 2016, from http://www.softschools.com/social_studies/state_abbreviations/